A Diagnostic Function for Bootstrap Confidence Intervals

Bradley Efron

efron@stanford.edu

Department of Statistics
Stanford University

The Standard Intervals

- Observe $\hat{\theta}$ as estimate of θ, standard error σ
- Approximate 95% two-sided interval for θ

$$\hat{\theta} \pm 1.96\sigma$$

- Exactly accurate and correct in a normal translation family

$$\hat{\theta} = \theta + \sigma Z, \qquad Z \sim \mathcal{N}(0,1)$$

- Automatic!

Student Score Data

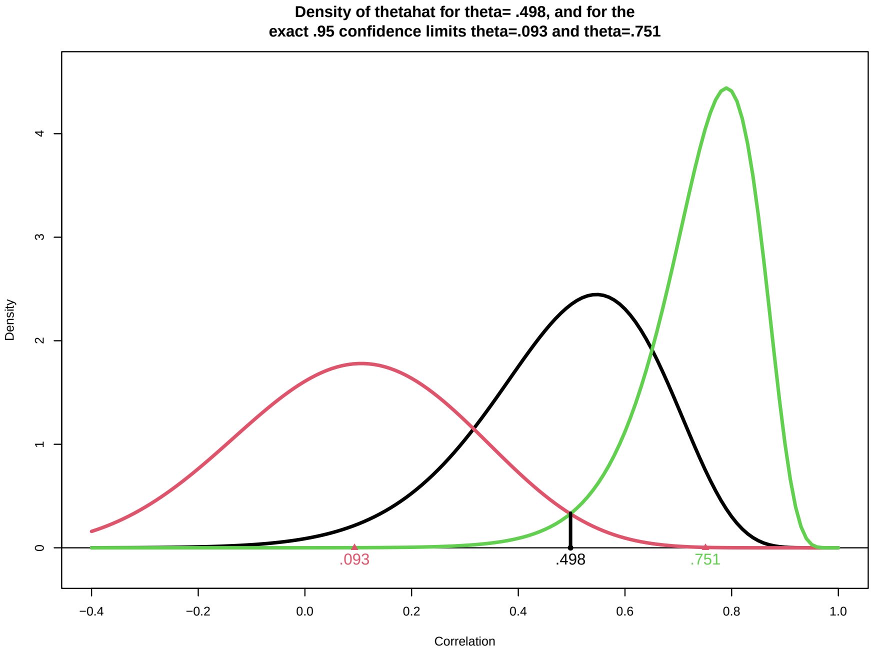
Mardia and Bibby

mechanics	vectors	algebra	analytics	statistics
7	51	43	17	22
44	69	53	53	53
49	41	61	49	64
59	70	68	62	56
34	42	50	47	29
⋮	⋮	⋮	⋮	⋮
42	69	61	55	45
46	49	53	59	37
63	63	65	70	63

Case Study: The Student Score Data

Mardia and Bibby

- $n = 22$ students have each taken five tests
- **Normal model** $\underset{22\times 5}{X}$ has independent rows $x_i \sim \mathcal{N}_5(\mu, \Sigma)$
- *Parameter of interest* $\theta = \text{correlation(mechanics,vectors)}$
- MLE $\hat{\theta} = 0.496$, $\hat{\sigma} = \frac{(1-\hat{\theta}^2)}{\sqrt{n-3}} = 0.173$
- Standard 95% interval $\theta \in (.160, .836)$



Density of thetahat for theta= .498, and for the
exact .95 confidence limits theta=.093 and theta=.751
Density
0
1
2
3
4
.093
.498
.751
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
Correlation

Fisher's Transformation Trick

- Monotone transformation $\quad \phi = m(\theta), \quad \hat{\phi} = m\left(\hat{\theta}\right)$

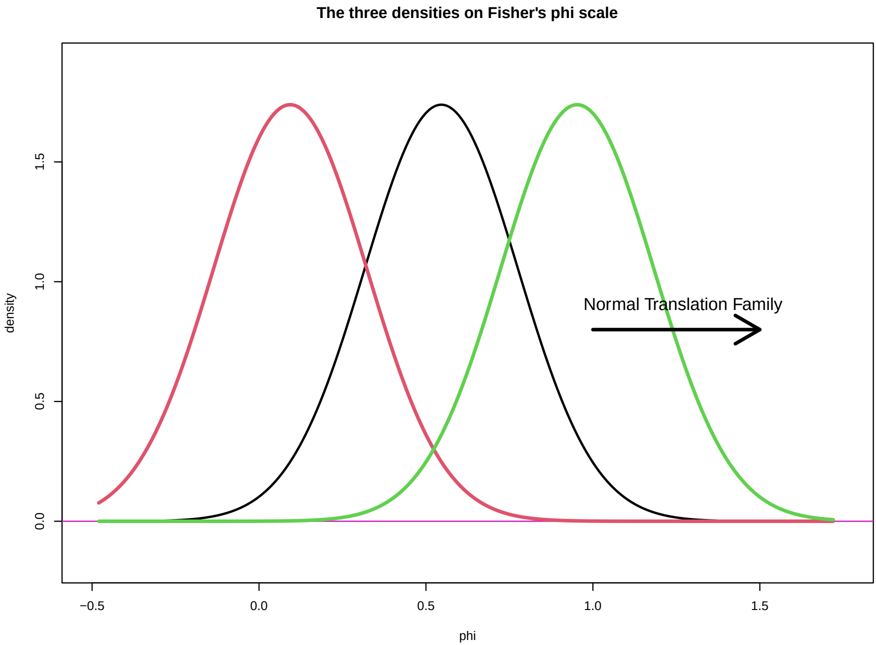
$$m(\theta) = \frac{1}{2}\log\left(\frac{1+\theta}{1-\theta}\right)$$

- **Normal translation family** $\quad \hat{\phi} \,\dot{\sim}\, \mathcal{N}(\phi, \sigma^2), \quad \sigma^2 = (n-3)^{-1}$

- Apply standard method on ϕ scale: $\hat{\phi} \pm 1.96\sigma$

- *Inverse transformation* $\quad \hat{\theta} = m^{-1}\left(\hat{\phi}\right)$ back to θ scale:

$$\theta \in m^{-1}\left(\hat{\phi} \pm 1.96\sigma\right)$$


The three densities on Fisher's phi scale
Normal Translation Family
density
phi
0.0
0.5
1.0
1.5
−0.5
0.0
0.5
1.0
1.5

Exact & Approximate Endpoints

for 95% interval

	.025	.975
Standard	.160	.836
Fisher	.097	.760
Exact	**.093**	**.751**
BCA	.086	.754

The Trouble With Fisher's Method

1. Requires Fisher
2. No nuisance parameters

- bca "bias-corrected and accelerated"
- A bootstrap algorithm for more accurate approximate confidence intervals
  - automatible — doesn't require a Fisher
  - allows for nuisance parameters

Maximum Eigenvalue Example

- Student score data matrix $\underset{22\times 5}{\boldsymbol{X}}$, assuming rows $x_i \overset{\text{ind}}{\sim} \mathcal{N}_5(\mu, \Sigma)$

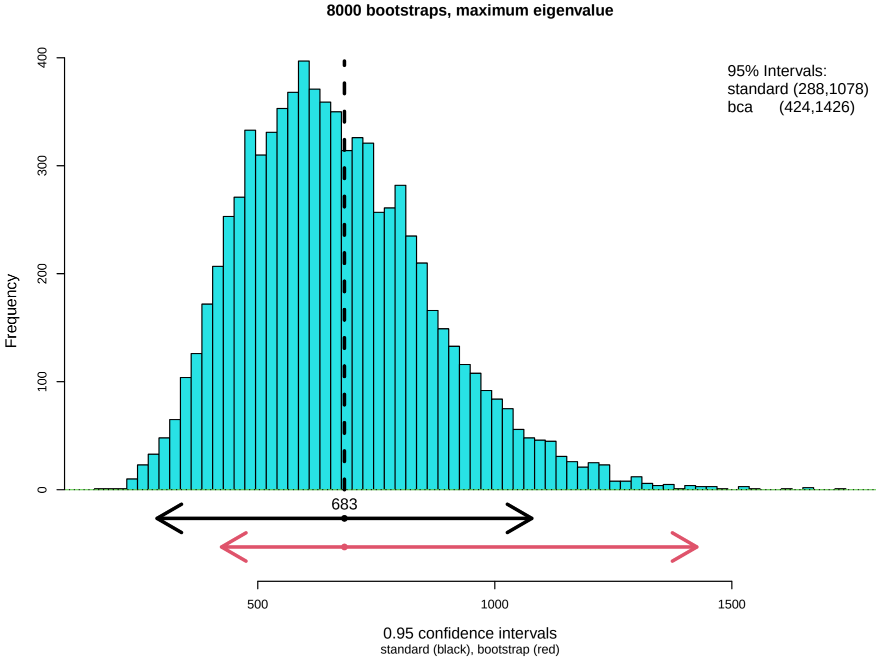
- Now let θ be the maximum eigenvalue of Σ, with MLE $\hat{\theta}$

$$\hat{\theta} = \text{maximum eigenvalue}\left(\widehat{\Sigma}\right) = \mathbf{683}$$

- $\mathcal{N}_5(\mu, \Sigma)$ is a 20-parameter family so there are 19 nuisance parameters

- **Parametric bootstrap** $x_i^* \overset{\text{ind}}{\sim} \mathcal{N}_5(\hat{\mu}, \widehat{\Sigma})$, $i = 1, \ldots, 22$

$$\Longrightarrow \underset{22\times 5}{\boldsymbol{X}^*} \Longrightarrow \widehat{\Sigma}^* \Longrightarrow \hat{\theta}^*$$


8000 bootstraps, maximum eigenvalue
95% Intervals:
standard (288,1078)
bca (424,1426)
Frequency
0
100
200
300
400
683
500
1000
1500
0.95 confidence intervals
standard (black), bootstrap (red)

Good Things About The BCA Intervals

- Automatable$\longrightarrow$ one program handles all cases
- Transformation invariant$\longrightarrow$ if $\phi = m(\theta)$ then $\hat{\phi}[\alpha] = m\left(\hat{\theta}[\alpha]\right)$ ($\hat{\theta}[\alpha]$ is bca level α upper endpoint)
- Second-order accurate$\longrightarrow$ actual coverage approaches nominal α at $O(n^{-1})$ compared to $O(n^{-1/2})$ for standard
- Correctness$\longrightarrow$ $\hat{\theta}[\alpha]$ is *correct* as well as accurate

NSTF: Normal Scaled Transformation Families

(Efron, 1982: *Ann Stat* 323–339)

- A one-parameter family of cdfs $\left\{F_\theta\left(\hat{\theta}\right)\right\}$ is NSTF if for some monotone transformation $\phi = m(\theta)$ and $\hat{\phi} = m\left(\hat{\theta}\right)$ we have

$$\hat{\phi} = \phi + \sigma_\phi (Z - z_0) \qquad [\sigma_\phi = 1 + a_0\phi,\ Z \sim \mathcal{N}(0,1)]$$

- z_0 "bias corrector"
- a_0 "acceleration"
- **Fisher** $\quad a_0 = z_0 = 0 \quad$ and $\quad m(\theta) = \frac{1}{2}\log\left(\frac{1+\theta}{1-\theta}\right)$

BCA Confidence Limits

(Efron, 1987: JASA 171–200)

- If $F_\theta\left(\hat{\theta}\right)$ is NSTF, the exact and correct one-sided upper level α confidence limit for θ is

$$\hat{\theta}[\alpha] = F_{\hat{\theta}}^{-1}\left\{\Phi\left(z_0 + \frac{z_0 + z^{(\alpha)}}{1 - a_0(z_0 + z^{(\alpha)})}\right)\right\}$$

- Don't need to know $m(\cdot)$!

- Gives standard interval if $a_0 = z_0 = 0$ and $F_\theta\left(\hat{\theta}\right) = \mathcal{N}(\theta, \sigma^2)$

Is The Family $\{F_\theta(\hat{\theta})\}$ An NSTF?

- For a given value of z let $\alpha = \Phi(z)$,

$$\hat{\theta}^{(\alpha)} = \alpha\text{th quantile of } \hat{\theta} \qquad \left(= F_\theta^{-1}(\alpha)\right)$$

$$\text{and} \quad \dot{F}_\theta\left(\hat{\theta}\right) = \frac{\partial}{\partial\theta} F_\theta\left(\hat{\theta}\right).$$

- The diagnostic function is

$$D_\theta(z) = \frac{\dot{F}_\theta\left(\hat{\theta}^{(\alpha)}\right) / \dot{F}_\theta\left(\hat{\theta}^{(0.5)}\right)}{\phi(z)/\phi(0)}$$

Theorem

$\left\{F_\theta\left(\hat{\theta}\right)\right\}$ is an NSTF iff $D_\theta(z)$ is linear in z.

Gamma Scale Family

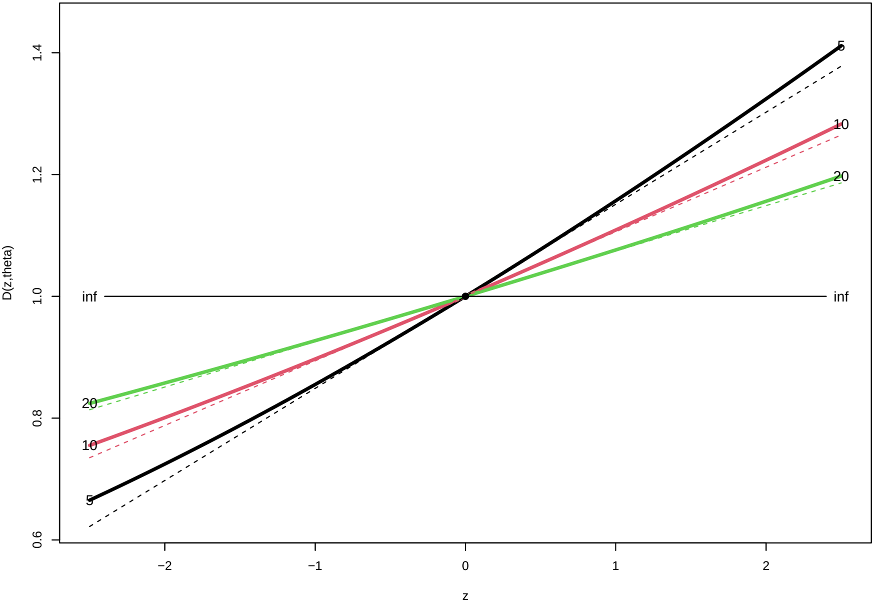
$$\hat{\theta} = \theta \cdot \text{Gamma}_\nu$$

$$f_\theta\left(\hat{\theta}\right) = \frac{\hat{\theta}^{\nu-1} e^{-\hat{\theta}/\theta}}{\theta^\nu \Gamma(\nu)} \qquad \text{for } \hat{\theta} \geq 0$$

$$F_\theta\left(\hat{\theta}\right) = \int_0^{\hat{\theta}} f_\theta(t)\, dt$$

- *Next* $\quad D_\theta(z) \quad$ (same for all θ)

**Diagnostic function D(z) for Gamma scale family,
df=5, 10, 20 and Infinity. Dashed lines are linear fits**



Student-t Translation Family

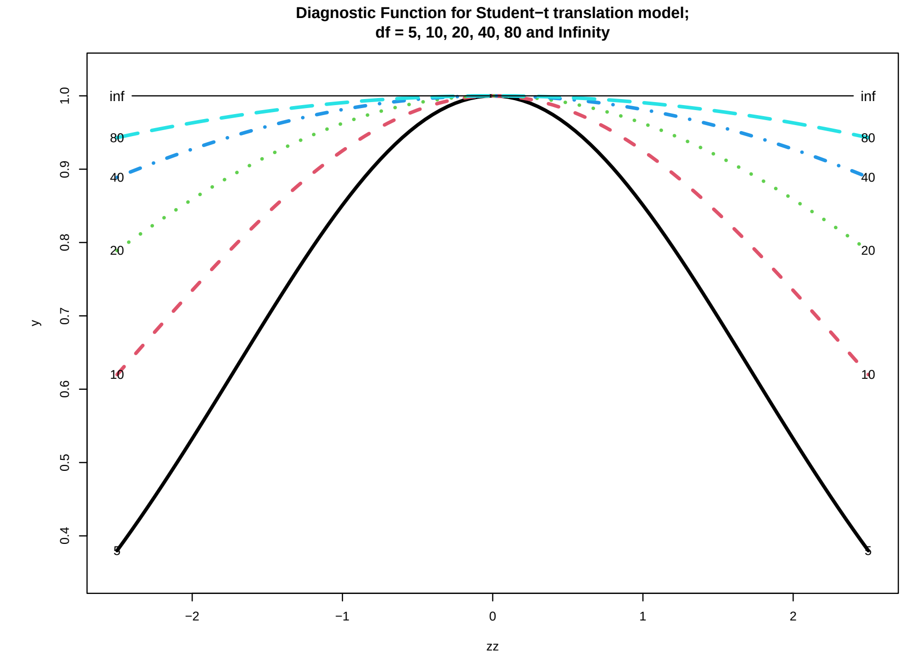
- One-parameter family

$$\hat{\theta} = \theta + W$$

where

W Student's t distribution with ν degrees of freedom

- *Next* $D_\theta(z)$ (same for all θ)


Diagnostic Function for Student−t translation model;
df = 5, 10, 20, 40, 80 and Infinity
inf
80
40
20
10
5
y
zz
−2
−1
0
1
2
0.4
0.5
0.6
0.7
0.8
0.9
1.0

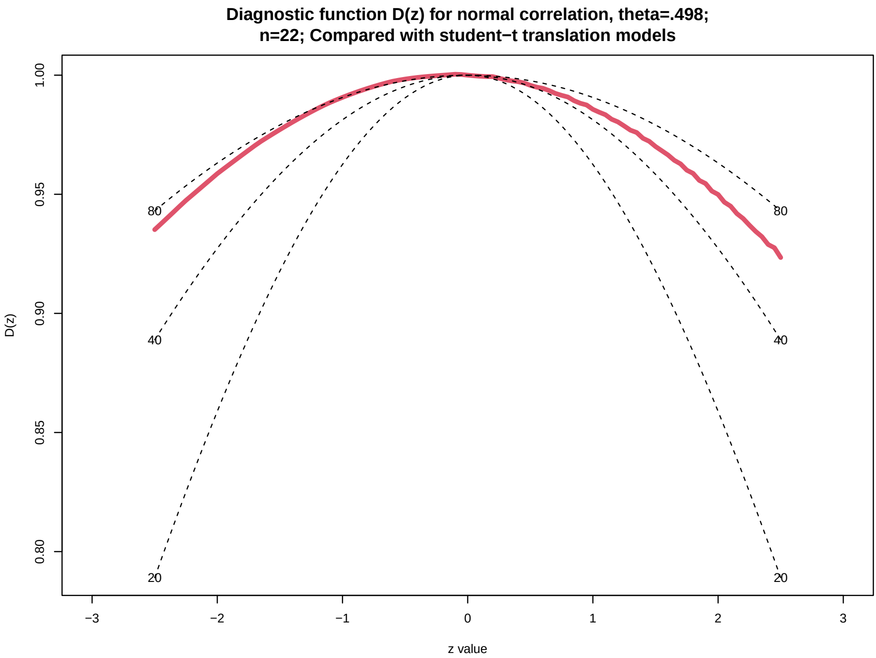
Normal Theory Correlation Coefficient

$$x_i \overset{\text{ind}}{\sim} \mathcal{N}_5(\mu, \Sigma), \qquad i = 1, \ldots, 22$$

$$\theta = \text{correlation}(X_1, X_2)$$

$$\hat{\theta} = \text{Pearson correlation coefficient}$$

- *Next* $D_{.498}(z)$


Diagnostic function D(z) for normal correlation, theta=.498;
n=22; Compared with student−t translation models
D(z)
1.00
0.95
0.90
0.85
0.80
80
40
20
80
40
20
−3
−2
−1
0
1
2
3
z value

GSTF: General Scaled Transformation Family

- **NSTF** $\quad \hat{\phi} = \phi + (1 + a_0\phi)(Z - z_0)$ with $Z \sim \mathcal{N}(0,1)$
- **GSTF** $\quad$ Replace Z with $W = w(Z)$
- $w(0) = 0$ and $w'(0) = 1$
- *Example* $\qquad W \sim \text{student}_\nu$ so $w(z) = T_\nu^{-1}(\Phi(z))$
- If $a_0 = z_0 = 0$: $\quad \hat{\phi} = \phi + W$

Diagnostic Function For A GSTF

Theorem

If $F_\theta\left(\hat{\theta}\right)$ is a GSTF then

$$D_\theta(z) = \frac{1 + w(z)a_0}{w'(z)}.$$

- For NSTF, $w(z) = z, w'(z) = 1$ and, for every value of θ,

$$D_\theta(z) = 1 + za_0,$$

linear in z, with slope a_0.

GSTF Confidence Limits

- **Fact:** For any choice of $w(\cdot)$ there's a version of the bca limits that is accurate and correct for that GSTF.
- *Idea:* Use estimated diagnostic function to estimate "w" and see if the GSTF limits are much different than the NSTF bca limit.

Multiparameter Models & Nuisance Parameters

- *Exponential families*: Observe data $\boldsymbol{X}$ with densities

$$f_\eta(\boldsymbol{X}) \propto e^{\eta^\top y} \qquad (\eta, y \in \mathcal{R}^p)$$

- $y = y(\boldsymbol{X})$ sufficient vector and $\eta =$ natural parameter vector
- $\mu = E_\eta\{y\}$ "expectation parameter"
- Parameter of interest

$$\theta = t(\mu), \qquad \text{with } \hat{\theta} = t(y) \text{ MLE}$$

- $p - 1$ nuisance parameters

Stein's Least Favorable Family

- Replace p-parameter family $f_\eta(\boldsymbol{X})$ with one-parameter family

$$\hat{f}_\lambda(\boldsymbol{X}) = f_{\hat{\eta}+\lambda \dot{\boldsymbol{t}}}(\boldsymbol{X}) \begin{cases} \hat{\eta} & = \text{MLE}\,\eta, \;\; \text{fixed} \\ \dot{\boldsymbol{t}} & = (\cdots \partial t(y)/\partial y_i \cdots) \end{cases}$$

- Back to one-parameter model

- Apply bca formula, diagnostic function, . . .

Student Score Maximum Eigenvalue Example

- Data $\underset{22\times 5}{\boldsymbol{X}}$ has rows $x_i \overset{\text{ind}}{\sim} \mathcal{N}_5(\mu, \Sigma)$, $i = 1, \ldots, 22$
- $y = (\bar{X}_1, \bar{X}_2 \ldots, \bar{X}_5, \bar{X}_1^2, \overline{X_1 X_2}, \ldots, \bar{X}_5^2)$
- $p = 20$
- $\theta =$ max eigenvalue (Σ) has MLE $\hat{\theta} =$ maxeig $(\widehat{\text{cov}}(\boldsymbol{X}))$

Calculating The BCA Intervals

(Efron and Narasimhan, 2020: CRAN package `bcaboot`)

- **Parametric bootstrap** $\hat{\eta}$ MLE of η,

$$f_{\hat{\eta}}(\cdot) \longrightarrow \boldsymbol{X}_1^*, \boldsymbol{X}_2^*, \ldots, \boldsymbol{X}_B^*$$

and

$$y_i^* = y(X_i^*), \quad \hat{\theta}_i^* = t(y_i^*) \qquad \text{for } i = 1, \ldots, B$$

- Program `bcapar` computes NSTF bca confidence limits; nonparametric version is `bcanon`

- Also gives diagnostic function $\widehat{D}(z)$: *linear*?

Estimated Diagnostic Function

- First-order Taylor approximation to $\hat{\theta}_i^*$ is

$$\hat{\theta}_i^* = \hat{\theta} + d_i^* \qquad \text{where}$$

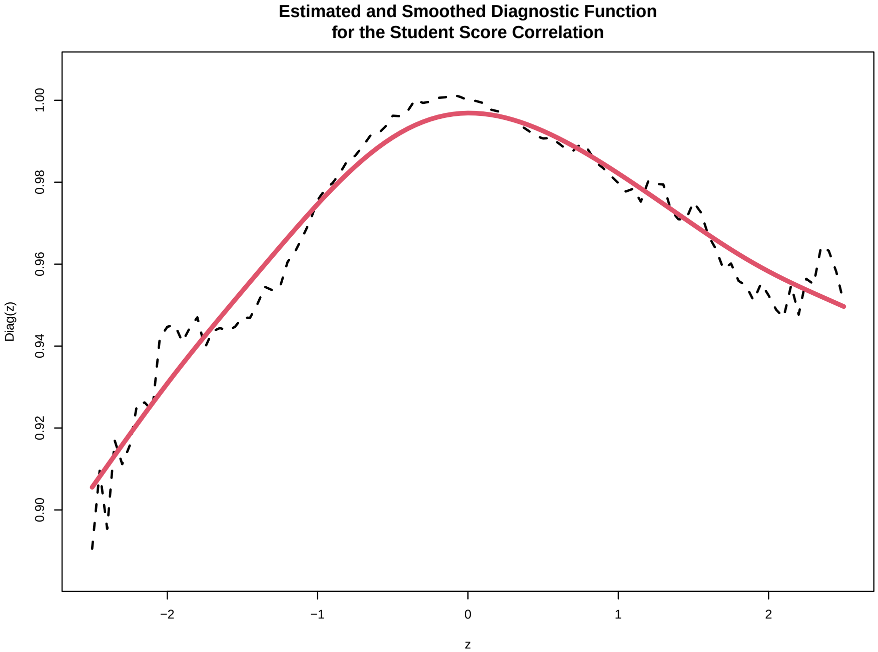
$$d_i^* = \boldsymbol{t}^\top (\boldsymbol{y}_i^* - \bar{\boldsymbol{y}}^*)$$

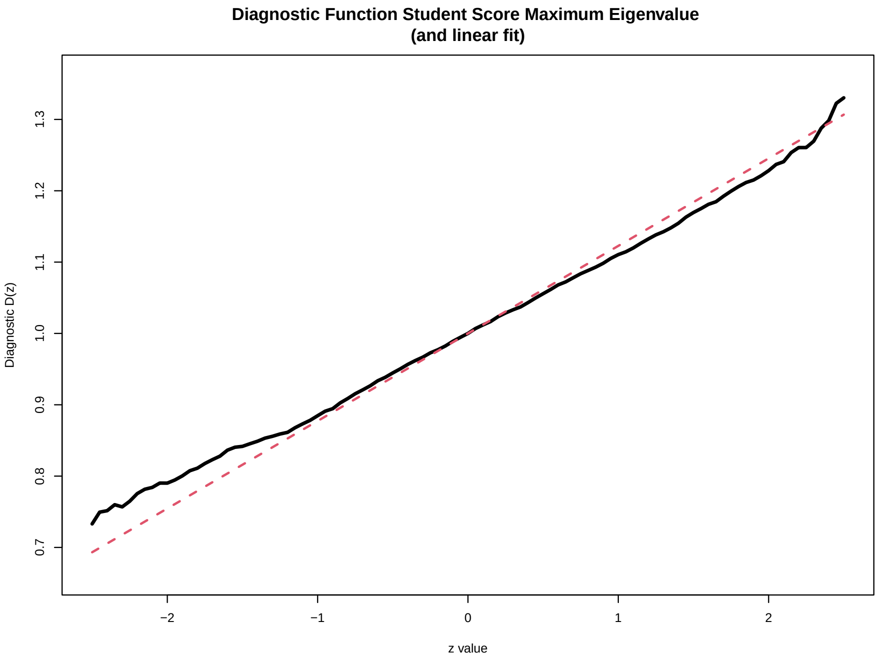
Lemma

$$\frac{\partial}{\partial \lambda} \widehat{F}_\lambda(z) \bigg|_{\lambda=0} = \frac{1}{B} \sum_{\hat{\theta}_i^* \leq \hat{\theta}^{*(\alpha)}} d_i^*$$

- $\alpha = \Phi(z)$ and $\hat{\theta}^{*(\alpha)}$ = boot αth quantile

- Gives diagnostic estimate $\widehat{D}_0(z)$


Estimated and Smoothed Diagnostic Function
for the Student Score Correlation
Diag(z)
z
0.90
0.92
0.94
0.96
0.98
1.00
−2
−1
0
1
2


Diagnostic Function Student Score Maximum Eigenvalue
(and linear fit)
Diagnostic D(z)
z value
0.7
0.8
0.9
1.0
1.1
1.2
1.3
−2
−1
0
1
2

Actual GSTF Levels

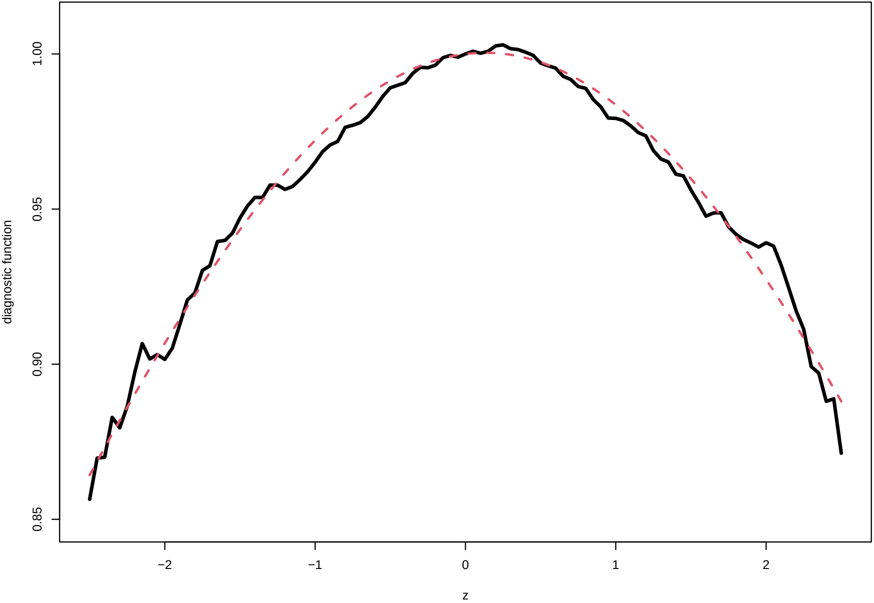
for nominal NSTF endpoints

	Student score maximum eigenvalue			
Nominal NSTF	.025	.050	.950	.975
Actual GSTF	.024	.049	.948	.975

Nonparametric BCA Confidence Intervals

- Student score data matrix $\underset{22\times 5}{\boldsymbol{X}}$ gives $\hat{\theta} = t(\boldsymbol{X})$
- **Bootstrap replication** $\underset{22\times 5}{\boldsymbol{X}^*}$: rows selected from $\boldsymbol{X}$ with replacement gives $\hat{\theta}^* = t(\boldsymbol{X}^*)$
- `bcanon(B,X,func)` gives bca confidence limits and diagnostic
- *Next* $B = 8000$ func $= \text{cor}(X_1, X_2)$

Nonparametric Diagnostic Function estimate, correlation example; Red is cubic fit; Close to student−t translation model df=44



Truth and Targets